

Bottom-up vs. Top-down methods for automatic profiling in Sweden

Barbro von Hofsten
Statistics Sweden



Outline of presentation

- History and recent work
- Definition of the statistical unit ENT
- Automatic profiling in Sweden
- The Bottom-up method
- The Top-down method
- Comparison between the different methods
- Remaining questions
- Conclusions

History of automatic profiling in Sweden

- Sweden part of the Eurostat Grants in 2015 (together with AT, BG, FI, FR, IT, LT, LV, NO, PL, SI) → The **first prototype for automatic profiling** was developed
- Swedish project on profiling in 2017-2018
 - ✓ Profiling strategy
 - ✓ Manual profiling
 - ✓ **Automatic profiling**
 - ✓ Consolidation
 - ✓ European profiling (grants)
 - ✓ LCU Portfolio
 - ✓ Implementation plan BR

Work so far in 2019

3 mini-projects focusing on:

- ✓ **Definitions of ENT** and KAU (institutional vs functional statistics)
- ✓ **Pros and cons with Top-down vs Bottom-up methods**
- ✓ Guidelines for manual profiling and visits
- ✓ Guidelines for light profiling/validation
- ✓ **Automatic profiling** and ancillary units
- ✓ Recommendations of a portfolio for the different types of profiling

Work so far in 2019 cont.

Impact assessment

- ✓ Business Register
- ✓ SBS, BD, FATS and 8 other surveys based on the statistical unit ENT

Definition of Enterprise unit according to Council Regulation (EEC) No 696/93

The enterprise is **the smallest combination** of legal units that is an organizational unit producing goods or services, which benefits from a **certain degree of autonomy in decision-making**, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit.

Status of automatic profiling in Sweden

- Only ad hoc development of the algorithm since the first (Bottom-up) prototype in 2015
- In 2019 first tests of a Top-down method for the total population (not only for manual profiling)
- No implementation of new statistical units in BR yet
- The SBS delivery in October 2019 will be the first big test for automatic profiling in Sweden
- It's urgent to make a decision on which model to use (Bottom-up or Top-down)!

Automatic profiling – core principles

- Direct ownership is an indicator of control
- LEUs with different sectors (financial/non-financial) can never end up in the same ENT
- Certain NACE codes are indicators of ancillary activity
 - > If a subsidiary is small and is typical ancillary (HQ, Real Estate, Wholesale, Logistics, Employment) we classify it as such.

The Bottom-up method

- The separate LEUs are the starting point (ENT=LEU)
- Grouping of legal units considering:
 - ✓ Level in enterprise group
 - ✓ Activity (LEUs with the same NACE are likely the same ENT)
 - ✓ Ancillary LEUs are grouped together with a primary producer
- Aggregation of establishments sets NACE on the ENT
- Focus on the first part of the legislative text: **The enterprise is the smallest combination of legal units that is an organizational unit producing goods or services**

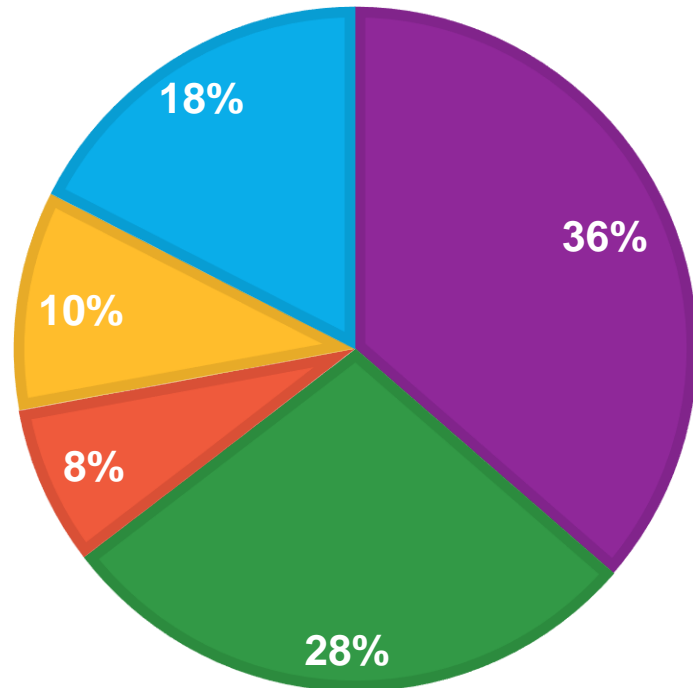
The Top-down method

- The Swedish part of the GEG is the starting point (ENT=Swedish GEG)
- LEUs in different sectors split the GEG into more than one ENT
- Aggregation of establishments sets NACE on the ENT
- Medium sized ENTs with indications of autonomous parts should be validated through light profiling
- Focus on the second part of the legislative text: The enterprise is the smallest combination of legal units that is an organizational unit producing goods or services, **which benefits from a certain degree of autonomy in decision-making** especially for the allocation of its current resources.

Comparison between the methods - Employment by degree of profiling

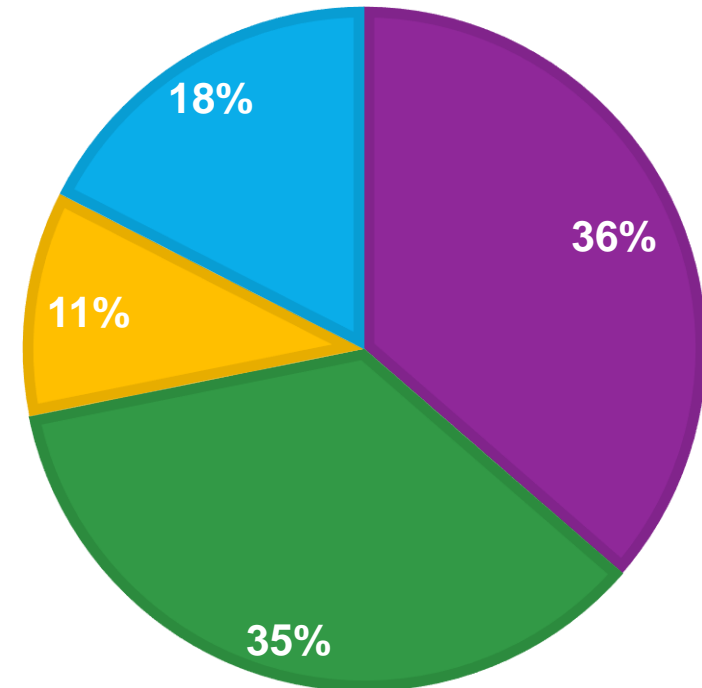
BOTTOM-UP

■ A. LEU ■ B. Automatic ■ C. ENT=LEU ■ D. Light/validation ■ E. Manual



TOP-DOWN

■ A. LEU ■ B. Automatic ■ D. Light/validation ■ E. Manual



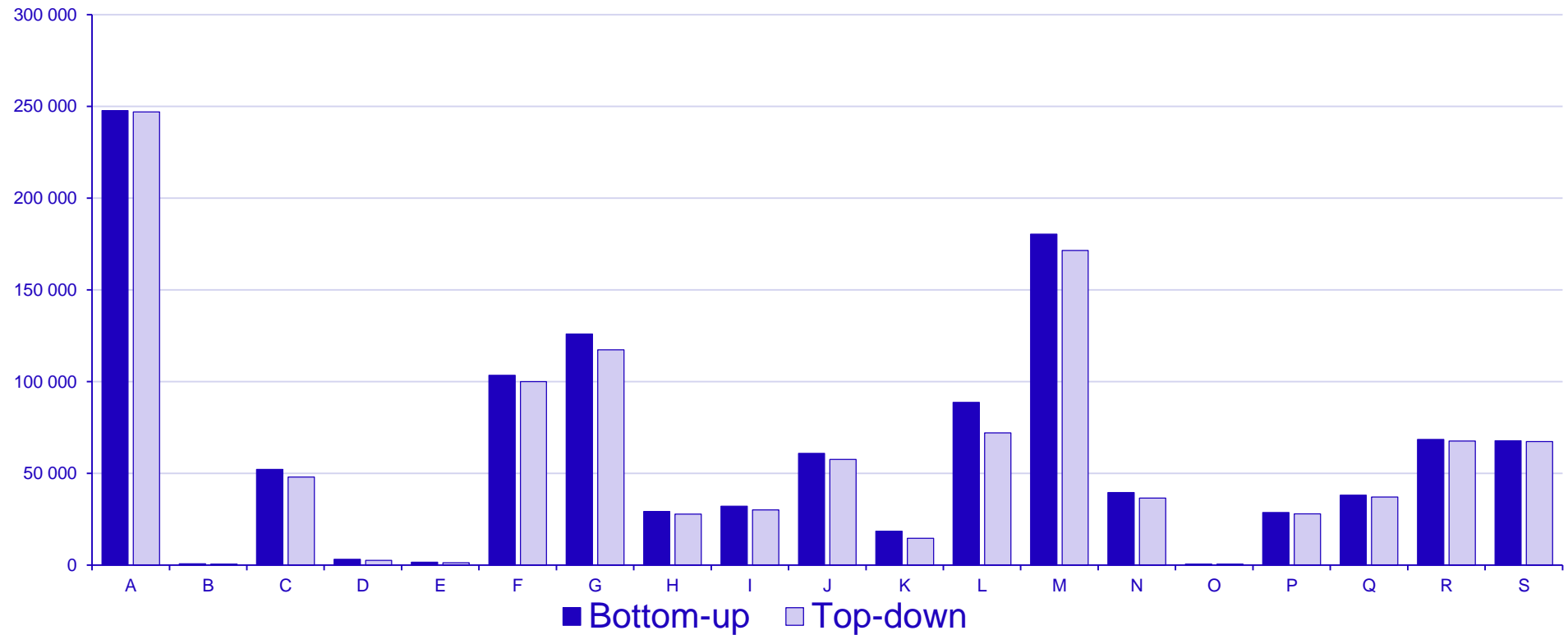
Number of ENTs in different size classes Bottom-up vs. Top-down (BR data)

Size class	Bottom-up	Top-down	Difference BU vs. TD
Micro enterprises (0-9)	1 151 937	1 095 373	56 564
Small enterprises (10-49)	37 056	32 574	4 482
Medium-sized enterprises (50-249)	6 533	5 515	1 018
Large enterprises (250-)	1 655	1 632	23
Total	1 197 181	1 135 094	62 087

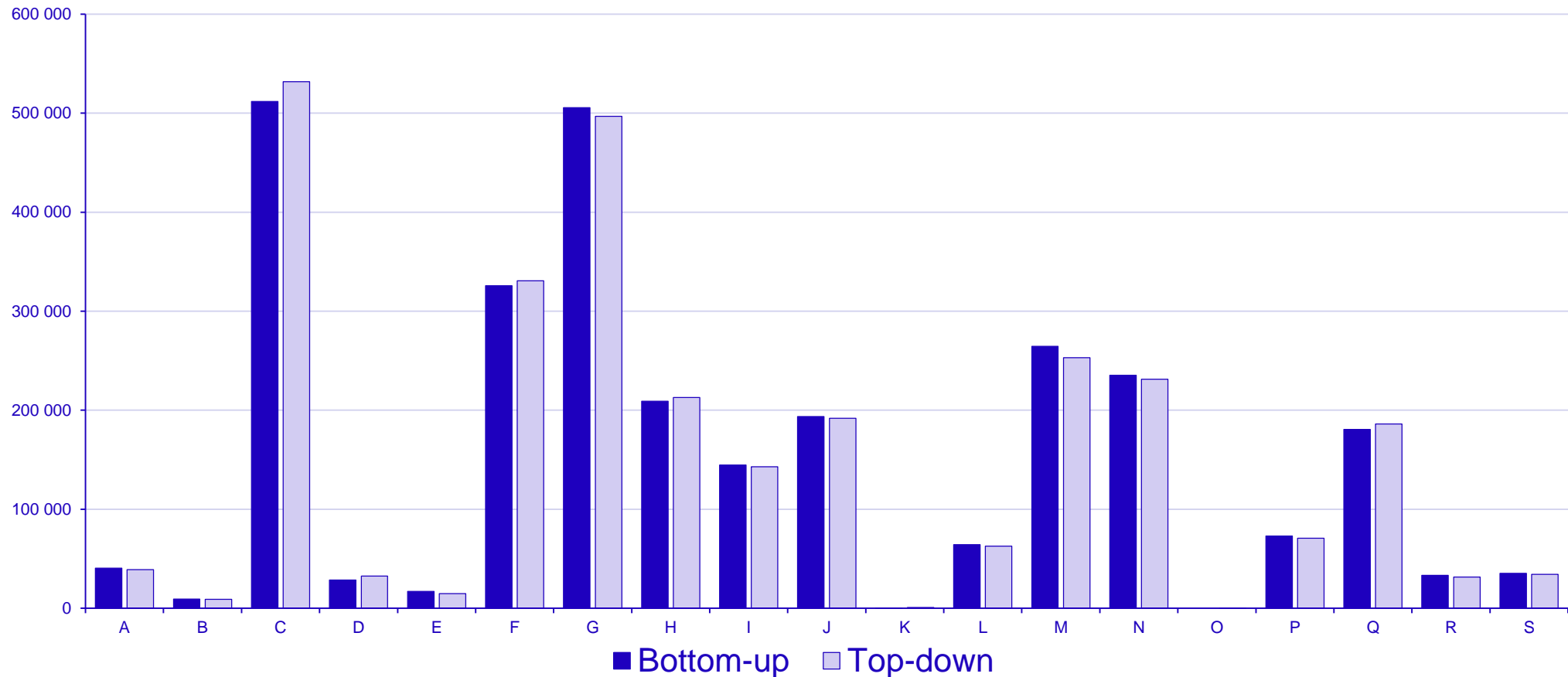
Number of employees (FTE) Bottom-up vs. Top-down (SBS data)

Size class	Bottom-up	Top-down	Difference BU vs. TD
Micro enterprises (0-9)	672 905	625 501	47 404
Small enterprises (10-49)	632 827	546 181	86 646
Medium-sized enterprises (50-249)	544 643	460 522	84 121
Large enterprises (250-)	1 021 807	1 240 058	-218 251

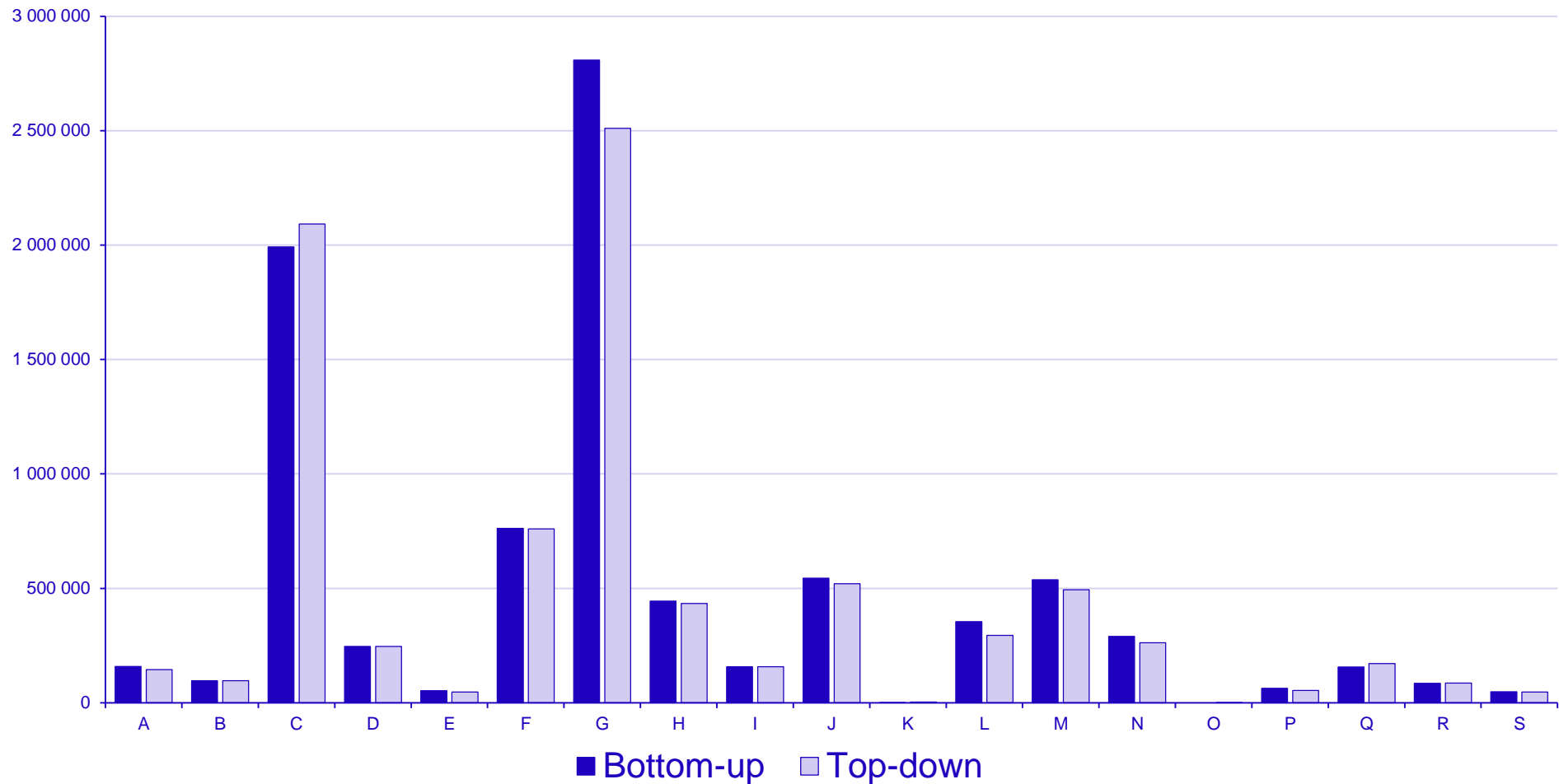
Number of ENTs in different NACE sections Bottom-up vs. Top-down (BR)



Number of FTE in different NACE sections Bottom-up vs. Top-down (SBS)



Consolidated turnover Bottom-up vs. Top-down (SBS)



Pros and cons with Bottom-up

Advantages:

- ✓ Possibility to get a step-by-step implementation in BR (smaller break in the time series)
- ✓ The bottom-up method is closer to the present situation

Disadvantages:

- ✓ A large amount of medium-sized enterprises will be left unattended
- ✓ Different methods for large and small enterprises
- ✓ Difficult to explain the term ENT to users and data providers

Pros and cons with Top-down

Advantages:

- ✓ The same approach can be used for the total population of enterprises (small, medium, large)
- ✓ Probably a better method for small and micro enterprises where $ENT=EG$ is a good assumption
- ✓ A simple algorithm compared to Bottom-up
- ✓ Creates the "Big-bang" requested from the users
- ✓ Easier to communicate to users and data providers

Disadvantages:

- ✓ Important to have enough resources for manual profiling and validation before the implementation.
- ✓ Maintenance of BR year $t+1$ will be more complicated

Remaining questions

- Legal aspects concerning data collection on ENT level
- Is it necessary to have a "Main-LEU" within each ENT?
- Practical BR-issues concerning ancillary units and sector code on ENTs
- Continuity rules for ENTs are a challenge. Continuity is very important for Business Demography
- Practical issues concerning data collection

Conclusions

- The Top-down method gives less numerous ENTs than Bottom-up
- The Top-down method probably gives a better distribution of the workforce in different size classes
- The Top-down method gives lower turnover figures than Bottom-up
- With the Top-down method there is a possibility to use the same profiling method for all enterprises
- The number of ENTs will not be perfect with any of the approaches (because the true value lies somewhere in between) 😊

Thank you for your attention!

Barbro von Hofsten

barbro.vonhofsten@scb.se

